P-N Junctions/Diodes

\[ V \]

\[ I \]

\[ \log I \]

slope = 60 mV/decade
No net current flow at thermal equilibrium:

\[
\begin{align*}
J_n &= q\mu_n \left( n\mathcal{E} + \frac{kT}{q} \frac{dn}{dx} \right) = 0 \\
\mathcal{E} &= -\frac{d\psi_i}{dx} \\
\frac{dE_f}{dx} &= 0 \\
n &= n_i e^{(E_f - E_i)/kT} = n_i e^{q(\psi_i - \psi_f)/kT}
\end{align*}
\]
Built-in Potential

- Fermi level $E_f$ is spatially constant (flat), causing a built-in potential difference across the diode.

- Built-in potential:

$$q\psi_{bi} = kT \ln \left( \frac{N_a N_d}{n_i^2} \right) = kT \ln \left( \frac{p_{p0}}{p_{n0}} \right) = kT \ln \left( \frac{n_{n0}}{n_{p0}} \right)$$

$\psi_{bi}$ = built-in potential
$p_p$ = (majority) hole density on p-side $\approx N_a$
$p_n$ = (minority) hole density on n-side
$n_n$ = (majority) electron density on n-side $\approx N_d$
$n_p$ = (minority) electron density on p-side
Abrupt Junctions: Depletion Approx.

Quasi-neutral p-region | Depletion region | Quasi-neutral n-region

\[ \begin{align*}
\rho(x) & = qN_d \\
E(x) & = -qN_a \\
\psi_i(x) - \psi_i(-x_p) & = \psi_m
\end{align*} \]

\[ \begin{align*}
& -\frac{d^2\psi_i}{dx^2} = \frac{qN_d}{\varepsilon_{si}} \quad \text{for } 0 \leq x \leq x_n \\
& -\frac{d^2\psi_i}{dx^2} = -\frac{qN_a}{\varepsilon_{si}} \quad \text{for } -x_p \leq x \leq 0 \\
E_m & = \left| \frac{-d\psi_i}{dx} \right|_{x=0} = \frac{qN_dx_n}{\varepsilon_{si}} = \frac{qN_ax_p}{\varepsilon_{si}} \\
\psi_m & = \frac{E_m(x_n + x_p)}{2} = \frac{E_mW_d}{2} \\
W_d & = \sqrt{\frac{2\varepsilon_{si}(N_a + N_d)\psi_m}{qN_aN_d}} \\
\psi_m & = \psi_{bi} \pm V_{app}
\end{align*} \]
One-Sided n⁺-p Diode

- Quasi-neutral n-region
- Depletion region
- Quasi-neutral p-region

Charge neutrality: \[ N_d x_n = N_a x_p, \] if \[ N_d \gg N_a \Rightarrow x_p \gg x_n, \]
i.e., depletion layer and voltage drop primarily appear on the lightly doped side.
One-Sided $n^+\text{-p}$ Diode

- Built-in potential: 
  \[ q \psi_{bi} \approx \frac{E_g}{2} + kT \ln \left( \frac{N_a}{n_i} \right) \]

\[ \Rightarrow \text{Built-in potential a weak function of doping conc.} \]
One-Sided n⁺-p Diode

- Depletion-layer width: \[ W_d = \sqrt{\frac{2\varepsilon_{si}(\psi_{bi} \pm V_{app})}{qN_a}} = x_n + x_p \approx x_p \]
- Depletion-layer capacitance: \[ C_d \equiv \frac{dQ_d}{dV_{app}} = \frac{\varepsilon_{si}}{W_d} \]
Quasi-Fermi Potentials $\phi_n$ and $\phi_p$

Nonequilibrium near the junction, $p_n \neq n_i^2$.

$$pn = n_i^2 \exp\left[q\left(\phi_p - \phi_n\right)/kT\right]$$

$$J_n = -q n \mu_n \left(\frac{d\psi_i}{dx} - \frac{kT}{qn} \frac{dn}{dx}\right) = -q n \mu_n \frac{d\phi_n}{dx}$$

$$J_p = -q p \mu_p \left(\frac{d\psi_i}{dx} + \frac{kT}{qp} \frac{dp}{dx}\right) = -q p \mu_p \frac{d\phi_p}{dx}$$

$$\phi_n = \psi_i - \frac{kT}{q} \ln\left(\frac{n}{n_i}\right)$$

$$\phi_p = \psi_i + \frac{kT}{q} \ln\left(\frac{p}{n_i}\right)$$
Spatial Variations of $\phi_n$ and $\phi_p$

$V_{app} = \phi_p - \phi_n$ at junction boundaries.

Inside the space-charge region: $J_n$ is constant (neglect G-R currents)

$$\Rightarrow \quad [n_n \mu_n d\phi_n/dx]_{xn} = [n_p \mu_p d\phi_p/dx]_{xp}$$

$$\Rightarrow \quad [d\phi_n/dx \text{ at } x_n] \ll [d\phi_p/dx \text{ at } -x_p]$$

$\Rightarrow \quad \phi_n \sim \text{constant inside space-charge region}$

**Practically all spatial variation in $\phi_n$ occurs in p-region,**

**Likewise, all spatial variation in $\phi_p$ occurs in n-region.**
Spatial Variations of $\phi_n$ and $\phi_p$

\[ V_{app} = \phi_p - \phi_n \] at junction boundaries.

\[ pn = n_i^2 \exp[q(\phi_p - \phi_n)/kT] \]

\[ V_{app} > 0 \] for forward bias, \[ V_{app} < 0 \] for reverse bias.

(1) \[ n_p(x = -x_p) \approx n_{p0}(x = -x_p) \exp(qV_{app}/kT) \]

(2) \[ p_n(x = x_n) \approx p_{n0}(x = x_n) \exp(qV_{app}/kT) \]

(1) and (2) are the most important boundary conditions governing a p-n diode.
Currents in a p-n Junction

- Generation-recombination currents in space-charge region are usually negligible.

⇒ Electron current leaving n-side = electron current entering p-side.
⇒ Hole current leaving p-side = hole current entering n-side.

- Need to consider minority carriers and currents only.

- Total current in diode = electron current + hole current.
Excess Electrons in the p-Region

\[ \frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} - R_n + G_n = 0 \]

\[ J_n = -q n \mu_n \left( \frac{d\psi}{dx} - \frac{kT n}{q n} \frac{dn}{dx} \right) = \mu_n kT \frac{dn}{dx} \]

\[ R_n - G_n = \frac{n - n_0}{\tau_n} \]

\[ \frac{d^2 n_p}{dx^2} - \frac{n_p - n_{p0}}{L_n^2} = 0, \]

where \( L_n = \sqrt{\tau_n D_n} = \sqrt{\frac{kT \mu_n \tau_n}{q}} \) is the minority carrier diffusion length.

Boundary conditions:
\( n_p = n_{p0} \exp\left( q V_{app} / kT \right) \)
at \( x=0 \),
and
\( n_p = n_{p0} \)
at \( x=W \) (ohmic contact).

\[ n_p - n_{p0} = n_{p0} \left[ \exp(q V_{app} / kT) - 1 \right] \frac{\sinh[(W - x) / L_n]}{\sinh(W / L_n)}. \]
Wide-Base and Narrow-Base Diodes

\[ J_n(x = 0) = qD_n \left( \frac{dn_p}{dx} \right) = -\frac{qD_n n_p 0 \left[ \exp(q V_{app} / kT) - 1 \right]}{L_n \tanh(W / L_n)} - \frac{qD_n n_i^2 \left[ \exp(q V_{app} / kT) - 1 \right]}{p_p 0 L_n \tanh(W / L_n)} \]

- **Wide-base: \( W >> L_n \)**
  - **Forward-bias:** \( J_n(x = 0) = -[qD_n n_i^2 / N_a L_n] \exp(q V_{app} / kT) \)
    - Current increases exponentially with \( V_{app} \), at 60 mV per decade at RT.
  - **Reverse-bias:** \( J_n(x = 0) = +[qD_n n_i^2 / N_a L_n] \)
    - Electrons on p-side but within a diffusion length of the depletion-region boundary diffuse towards n-side.

- **Narrow-base: \( W << L_n \)**
  - **Forward-bias:** \( J_n(x = 0) = -[qD_n n_i^2 / N_a W] \exp(q V_{app} / kT) \)
  - **Reverse-bias:** \( J_n(x = 0) = +[qD_n n_i^2 / N_a W] \)
  - **Currents increase rapidly as \( W \) decreases!**
Most p-n junctions in silicon technology are narrow-base diodes. (e.g., source and drain in MOSFETs, emitter and base in bipolars.)
Turning Off a p-n Diode

- Excess electrons in p-region
  \[ Q_B = -q \int_{0}^{W} (n_p - n_{p0}) \, dx \]
- Effective turn off starts only after most excess minority carriers have recombined or have drained off.
Diffusion Capacitance

- Diffusion capacitance $C_D$ is due to stored minority carriers responding to applied voltage.

- $C_D$ due to electrons stored in p-type region:
  \[ C_{Dn} = \frac{dQ_B}{dV_{app}} \propto \exp\left(\frac{qV_{app}}{kT}\right) \]

- For a diode or bipolar transistor to switch fast, it must have minimal diffusion capacitance.

- To minimize diffusion capacitance: increase doping concentration and minimize charge-storage volume.

- Modern high-speed bipolar transistors require very thin base.
MOS Device

Gate electrode (metal or polysilicon)

Vacuum level

$q\phi_m = 4.10\ eV$

$E_f$

Metal (aluminum)

Vacuum level

$E_g = 1.12\ eV$

$E_i$

Silicon (p-type)

$\phi_s = \chi + \frac{E_g}{2q} + \psi_B$

Silicon dioxide

Vacuum level

$0.95\ eV$

$E_c$

$8-9\ eV$

$E_v$

Silicon dioxide

$E_f$

$E_g$

$q\chi = 4.05\ eV$

$q\phi_s$

$E_c$

$q\psi_B$
Assume $\phi_m = \phi_s$.
Poisson’s Equation

\[
\frac{d^2 \psi}{dx^2} = -\frac{d \mathcal{E}}{dx} = -\frac{q}{\varepsilon_{si}} \left[ p(x) - n(x) + N_d^+ (x) - N_a^- (x) \right]
\]

\[
p(x) = n_i e^{q(\psi_f - \psi_i)/kT} = n_i e^{q(\psi_a - \psi)/kT} = N_a e^{-q\psi/kT}
\]

\[
n(x) = n_i e^{q(\psi_i - \psi_f)/kT} = n_i e^{q(\psi - \psi_a)/kT} = \frac{n_i^2}{N_a} e^{q\psi/kT}
\]
Solving Poisson’s Equation

\[ \varepsilon^2(x) = \left( \frac{d\psi}{dx} \right)^2 = \frac{2kTN_a}{\varepsilon_{si}} \left[ \left( e^{-\frac{q\psi}{kT}} + \frac{q\psi}{kT} - 1 \right) + \frac{n_i^2}{N_a^2} \left( e^{\frac{q\psi}{kT}} - \frac{q\psi}{kT} - 1 \right) \right] \]

\[ Q_s = -\varepsilon_{si} \varepsilon_s = \pm \sqrt{2\varepsilon_{si} kT N_a} \left[ \left( e^{-\frac{q\psi_s}{kT}} + \frac{q\psi_s}{kT} - 1 \right) + \frac{n_i^2}{N_a^2} \left( e^{\frac{q\psi_s}{kT}} - \frac{q\psi_s}{kT} - 1 \right) \right]^{1/2} \]
Depletion Approximation
(1-D Uniform Doping)

\[ \begin{align*}
Q_d &= qN_a W_d \\
\varepsilon &= qN_a (W_d - x) / \varepsilon_{si} \\
\psi &= qN_a (W_d - x)^2 / 2 \varepsilon_{si} \\
\Rightarrow \psi_s &= qN_a W_d^2 / 2 \varepsilon_{si} \\
W_d &= \sqrt{\frac{2 \varepsilon_{si} \psi_s}{qN_a}} \\
\psi &= \psi_s \left(1 - \frac{x}{W_d}\right)^2
\end{align*} \]
Condition for Strong Inversion

\[ \psi_s (\text{inv}) = 2\psi_B = 2 \frac{kT}{q} \ln \left( \frac{N_a}{n_i} \right) \]

i.e., \((n_i^2/N_a^2)\exp(q\psi_s/kT) = 1.\)

And the electron concentration at the surface equals the hole concentration in the bulk Si.
In contrast to p-n junctions, $W_d$ reaches a maximum value $W_{dm}$ at the onset of strong inversion when

$$\psi_s = 2\psi_B = 2(kT/q)\ln(N_a/n_i):$$

$$W_{dm} = \sqrt{\frac{4\varepsilon_{si}kT\ln(N_a/n_i)}{q^2N_a}}$$

This defines the threshold condition of a MOSFET. $W_{dm}$ also plays a key role in the short-channel scaling of a MOSFET, namely, $L_{min} \propto W_{dm}$. 

![Graph showing the relationship between substrate doping concentration and maximum depletion width](image)
Strong Inversion

\[
\frac{d\psi}{dx} = -\sqrt{\frac{2kTN_a}{\varepsilon_{si}} \left( \frac{q\psi}{kT} + \frac{n_i^2}{N_a^2} e^{q\psi/kT} \right)}
\]

Charge per area:
\[
Q_i = -\sqrt{\frac{2\varepsilon_{si} kT n_i^2}{N_a}} e^{q\psi_s/2kT}
\]

Electron conc. at surface:
\[
n(0) = \frac{n_i^2}{N_a} e^{q\psi_s/kT}
\]

Inversion layer thickness:
\[
\frac{Q_i}{qn(0)} = 2\varepsilon_{si} kT / qQ_i
\]

Electron concentration, \(n(x)\) (cm\(^{-3}\))

Distance from surface, \(x\) (Å)

\(N_a = 10^{16}\) cm\(^{-3}\)

\(\psi_s = 0.88\) V

\(\psi_s = 0.85\) V
Quantum Effect in MOS Inversion

In an MOS inversion layer, carriers are confined in the direction perpendicular to the surface and therefore need to be treated quantum mechanically (2-D).

Discrete energy levels:

\[ E_j = \left[ \frac{3hqE_s}{4\sqrt{2m_x}} \left( j + \frac{3}{4} \right) \right]^{2/3} \]

Average distance of inversion layer from the surface:

\[ x_j = \frac{2E_j}{3qE_s} \]
Electron ground state is at some finite energy above the bottom of the conduction band.

Band bending must exceed $2\psi_B$ to invert surface.

The centroid of inversion layer is farther away from the surface than in the classical case.
MOSFET Charge and Potential

Gate voltage equation ($V_{fb} = 0$):

$$V_g = V_{ox} + \psi_s = -\frac{Q_s}{C_{ox}} + \psi_s$$

$$Q_s = -\varepsilon_s \varepsilon_{ox} = -\varepsilon_{si} \varepsilon_{ox} = \varepsilon_{si} \varepsilon_s$$

Note: $C_{ox} = \varepsilon_{ox} / t_{ox}$

and $\varepsilon_{ox} \varepsilon_{ox} = \varepsilon_{si} \varepsilon_s$

$N_a = 10^{17}$ cm$^{-3}$  
$t_{ox} = 10$ nm
Inversion Charge in Log Scale

![Graph showing the relationship between gate voltage and inversion charge density with labels for \( Q_i \), \( 2\psi_B \), \( N_a = 10^{17} \text{ cm}^{-3} \), and \( t_{ox} = 10 \text{ nm} \).]
MOS Capacitances

\[ C_{ox} \]

\[ C_{si} \]

\[ \frac{1}{C} = \frac{1}{C_{ox}} + \frac{d\psi_s}{d(-Q_s)} \]

\[ C_{si} = \frac{d(-Q_s)}{d\psi_s} \]

\[ C = \frac{d(-Q_s)}{dV_g} \]

\[ V_g \]

\[ Q_s \]

\[ Q_d \]

\[ Q_i \]

\[ \psi_s \]
In accumulation, $Q_s \propto \exp(-q\psi_s/2kT)$, so

$$C_{si} = -\frac{dQ_s}{d\psi_s} = \left(\frac{q}{2kT}\right)Q_s = \left(\frac{q}{2kT}\right)C_{ox}|V_g - \psi_s|.$$
At flatband voltage, \( \frac{q \psi_s}{kT} \ll 1 \), therefore, \( Q_s = -\left(\varepsilon_{si} q^2 N_a / kT\right)^{1/2} \psi_s \).

\[
\frac{1}{C_{fb}} = \frac{1}{C_{ox}} + \sqrt{\frac{kT}{\varepsilon_{si} q^2 N_a}} = \frac{1}{C_{ox}} + \frac{L_D}{\varepsilon_{si}}
\]
Capacitance-Voltage Characteristics

- In depletion,

\[
\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_d}
\]

where

\[
C_d = \left| \frac{d(-Q_d)}{d\psi_s} \right| = \sqrt{\frac{\varepsilon_{si}qN}{2\psi_s}} = \frac{\varepsilon_{si}}{W_d}
\]

Note that

\[
V_g = \frac{qN_a W_d}{C_{ox}} + \psi_s = \sqrt{2\varepsilon_{si}qN_a \psi_s} + \psi_s
\]
Capacitance-Voltage Characteristics

- Inversion, high freq.: Inversion charge cannot respond,
  \[
  \frac{1}{C_{\text{min}}} = \frac{1}{C_{ox}} + \sqrt{\frac{4kT \ln(N_a/n_i)}{\varepsilon_s q^2 N_a}}
  \]

- Inversion, low freq., or connected to a reservoir:
  \[
  \frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_d + C_i}
  \]
  where
  \[
  C_i = \frac{d(-Q_i)}{d\psi_s} = \frac{|Q_i|}{2kT/q}
  \]
  is the inv. layer cap.

Like accumulation, \[
\frac{1}{C} = \frac{1}{C_{ox}} \left[ 1 + \frac{2kT}{q \left| V_g - \psi_s \right|} \right]
\]
Split C-V Measurement
Effect of Gate Work Function

\[ V_t = V_{fb} + 2\psi_B + V_{ox} = V_{fb} + 2\psi_B + \frac{Q_d}{C_{ox}} \]

\[ V_{fb} = (\phi_m - \phi_s) - \frac{Q_{ox}}{C_{ox}} \]

\[ \phi_s = \chi + \frac{E_g}{2q} + \psi_B \]

\[ \phi_m = \chi \begin{cases} (n^+ \text{ poly}) \\ \frac{E_g}{2q} \quad \text{(midgap)} \\ \frac{E_g}{q} \quad \text{(p\textsuperscript{+} poly)} \end{cases} \]
Effect of Gate Work Function

Example: n⁺ polysilicon gate on p-type silicon

\[
\phi_{ms} = -\frac{E_g}{2q} - \psi_B = -0.56 - \frac{kT}{q} \ln \left( \frac{N_a}{n_i} \right)
\]

\[V_g = 0\]

\[V_g = V_{fb} = \phi_{ms}\]
Poly-Si Gate Depletion Effect

n+ poly Oxide p-type silicon

Gate eq. becomes:

\[ V_g = V_{fb} + \psi_s + \psi_p - \frac{Q_s}{C_{ox}} \]

and,

\[ \frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_{si}} + \frac{1}{C_p} \]

Typically, \( t_{inv} \) is 0.8-1.0 nm thicker than \( t_{ox} \).
Gated-Diode: MOS + p-n Junction

Zero-bias on the p-n junction (equilibrium):

The electron quasi-Fermi level in the MOS is the same as the Fermi level of the p-type Si.

⇒ Inversion occurs when $\psi_s = 2\psi_B$. 
Gated-Diode: Reverse Biased (Nonequilibrium)
Spatial Variations of $\phi_n$ and $\phi_p$

\[ V_{app} = \phi_p - \phi_n \] at junction boundaries.

\[ pn = n_i^2 \exp[q(\phi_p - \phi_n)/kT] \]

$V_{app} > 0$ for forward bias, $V_{app} < 0$ for reverse bias.

(1) \[ n_p(x = -x_p) \approx n_{p0}(x = -x_p) \exp(qV_{app}/kT) \]

(2) \[ p_n(x = x_n) \approx p_{n0}(x = x_n) \exp(qV_{app}/kT) \]

(1) and (2) are the most important boundary conditions governing a p-n diode.
MOS under Nonequilibrium

For a p-n junction reverse-biased at a voltage $V_R$, the electron concentration on the p-side of the junction is

$$n = \frac{n_i^2}{N_a} e^{-qV_R/kT}$$

If a gate voltage is applied to bend the p-type bands by $\psi_s$, the electron concentration at the surface is

$$n = \frac{n_i^2}{N_a} e^{q\psi_s/kT} e^{-qV_R/kT}$$

For surface inversion to occur, i.e., $n = N_a$, Need

$$\psi_s(\text{inv}) = V_R + 2\psi_B$$
MOS under Nonequilibrium

Maximum depletion width at inversion is

\[ W_{dm} = \frac{\sqrt{2\varepsilon_{si} (V_R + 2\psi_B)}}{qN_a} \]